

# 10分布

2019年2月19日 17:15



## ◆ 分布式文件系统

1. 本地文件系统Local File System: 仅允许本地CPU通过IO总线访问存储设备
2. 分布式文件系统Distributed File System: 建立在松散耦合MPS上的, 通过通信网络或计算机网络互连, 将分布在若干节点的存储设备以共享文件系统方式统一管理, 提供给不同节点上的用户共享

1) 基于客户机/服务器模式, 是相对独立的软件系统, 被集成在分布式系统

### 一. 分布式系统distributed system

- 1) 定义: 是基于软件实现的一种MPS, 多处理机通过通信线路互连而构成松散耦合系统, 系统的处理和控制在各处理机上
- 2) 或者说, 是利用软件系统构建在计算机网络上的一种MPS
- 3) 又或者, 分布式系统是独立计算机的集合, 用户体验却像一台计算机
- 4) 又或者, 是能为用户自动管理资源的网络操作系统, 负责调用完成用户任务所需资源, 整个网络像一个大的计算机系统一样, 对用户透明
- 5) 与前述的MPS区别有
  - (1) 各节点都是独立计算机, 配有完整的外部设备
  - (2) 节点耦合程度更分散, 地理分布区域更广阔
  - (3) 各节点都能运行不同的操作系统, 有各自的文件系统, 除了本节点的管理外, 还有其他多个机构对其实施管理

### 1. 分布式系统的特征

- 1) 分布性
  - (1) 计算机地理位置分布性: 位置和地域范围分散且广阔
  - (2) 系统功能的分布性: 分散在各节点计算机上
  - (3) 系统资源分布性/自治性: 分散配置在各节点计算机上
  - (4) 系统控制的分布性/自治性: 各计算机一般无主从之分
- 2) 透明性: CPU、文件、打印机等, 被共享, 用户可使用其他节点的资源
- 3) 同一性: 程序可分布在多台计算机上并行运行, 即计算机可协作完成任务
- 4) 全局性: 系统具备全局进程通信机制, 任两台计算机可交换信息

### 2. 分布式系统的优点

- (1) 缺点是专用软件少, 网络安全的潜在不足多
- 1) 计算能力强: 并行处理技术, 使任务可被划分、分配给多节点并行执行
- 2) 易实现共享: 基于计算机网络使资源易共享, 计算迁移功能使负载易共享
- 3) 方便通信: 基于计算机网络和地域范围广阔为信息交流提供了方便
- 4) 可靠性高: 工作负载分散使单个机器故障不影响其他机器
- 5) 可扩充性好: 不需修改系统软硬件即可添加若干台计算机

### 3. 分布式操作系统

- 1) 是配置在分布式系统上的公用操作系统, 以全局的方式对分布式系统中的

所有资源进行统一管理，直接统一管理系统中地理位置分散的各物理和逻辑资源，有效协调和控制各任务的并行执行，协调和保持系统内的信息传输及协作运行，向用户提供统一的、方便的、透明的使用系统的界面和标准接口。典型的例子是world wide web及其唯一操作页面：web页面

- 2) 与网络操作系统不同，分布式系统用户使用系统内资源时不需了解各计算机的功能与配置、操作系统的差异、软件资源、网络文件的结构、物理设备的地址、远程访问的方式等，系统内部实现的细节也对用户屏蔽
- 3) 保持了网络操作系统功能的同时具有透明、内聚、可靠、高性能等特点
- 4) 除单机系统具有的功能以外还应实现
  - (1) 通信管理：提供通信机制和方法，使不同结点的进程能方便地交换信息。一般是通过提供符合通信协议的原语实现通信
  - (2) 资源管理：统一管理、分配、调度所有系统资源，方便用户使用，提高资源利用率。如分布式文件系统，分布式数据库系统、分布式程序设计语言及编译系统、分布式邮件系统等
  - (3) 进程管理：提供进程或计算迁移，实现平衡结点负载，加速计算速度；提供分布式同步和互斥机制，实现协调资源共享和竞争，提供进程并行程度，应对死锁

## 二. 分布式文件系统的实现方式和基本要求

### 1. DFS的实现方式

- 1) 共享文件系统方式shared file system approach/专用服务器方式
  - (1) 类似本地文件系统的树形目录结构，用逻辑树结构管理整个系统
  - (2) 客户/服务器模式，设置若干的文件服务器，数据分布其中
  - (3) 用户无需关注物理位置，只需按逻辑关系，即可通过文件服务器的服务进程，访问整个系统的共享资源
  - (4) 服务器分布和逻辑树架构对用户透明。如NFS、AFS、Sprite等系统
  - (5) 实现简单，设备要求低，通用性好
- 2) 共享磁盘方式shared disk approach/无服务器方式
  - (1) 配置一共享磁盘，与主机、客户机都连接在内部高速网络（如光通道上），主机和客户机都将共享磁盘视作其存储设备，以盘块方式读写磁盘上的文件，实现共享，如VAX Cluster、IBM GPFS和GFS
  - (2) 设备要求高，实现难度大，一般用来构造高端货专用存储设备如Network Attached Storage和Storage Area Network

### 2. 基本要求

- (1) 相比LFS，DFS除了大容量外，还要求
  - 1) 透明性：以下特点对用户透明
    - (1) 位置：文件服务器及文件服务进程的多重性，共享存储器的分散性
    - (2) 移动：存储资源 and 数据在系统内位置的移动
    - (3) 性能：系统中的服务负载有一定范围的变化量，保障性能稳定
    - (4) 扩展：在规模、性能和功能上允许系统扩展以满足负载和网络规模
  - 2) 高性能和高可靠性，满足诸多客户繁重的访问需求，保证系统安全可靠

- 3) 容错性：保证信息可靠性，系统出错时仍能为用户提供服务，一般通过冗余技术实现故障的掩盖：信息冗余，如增加校验信息；时间冗余，如重复执行操作；物理冗余，如增加额外的副本、设备或进程等
- 4) 安全性，通过身份验证、访问控制和安全通道等
- 5) 一致性，保证客户在本地缓存的文件副本与文件服务器上的一致

### 三. 命名及共享语义

1. 命名：除了LFS的屏蔽盘面、磁道号、扇区的抽象外，还需隐藏服务器地址和存储方式等细节，提供多级映射的抽象访问名接口
  - 1) 结合主机名和本地名，简单易行但不透明，如/host/local-name-path
  - 2) 将若干服务器的远程目录加载到客户机的本地目录。复杂度高，结构混乱，安全度低，一个服务器故障会导致客户机上的目录集失效
  - 3) 全局统一命名，每个文件和目录使用唯一名。考虑不到特殊文件，难实现
2. 共享语义
  - (1) 为保证多客户机并发访问时的数据一致性，必须精确处理客户机和服务器之间的交互协议，即精确定义读和写
    - 1) UNIX语义：文件的每个操作对所有进程都是瞬间可见
    - 2) 会话语义：文件关闭前，所有改动对进程都不可见
    - 3) 不允许更新文件，只能简单共享和复制
    - 4) 事务处理：所有改动以原子操作的方式顺序发生
3. 租赁协议
  - 1) 是一种有代表性的一致性访问协议，是一种多读者单写者机制，收到待读数据的同时会收到一个租赁凭据，上面附带保证数据不会被更新的有效期
  - 2) 在该有效期内读数据可直接访问在本地缓存的副本，过期后需要向服务器确认是否发生了更新，是否要接受新数据
  - 3) 服务器收到了更新请求时，需向所有该数据的租赁客户机发出更新确认，收到更新确认的客户机立刻标记凭据过期

### 四. 远程文件访问和缓存

- 1) 根据程序的局部性原理，DFS调用远程过程调用RPC，送回给本地缓存的数据量会比实际请求的多得多
1. 缓存和远程服务的比较
  - 1) 将远程访问转为本地缓存访问，加快了访问速度
  - 2) 使用本地缓存，减少了服务器负载和网络通信量，加强了扩充能力
  - 3) 使用本地缓存，一般减少了网络总开销
  - 4) 使用缓存带来了一致性问题，频繁写的情况下使上述方面优势转为劣势
  - 5) 使用本地缓存使仿真集中式系统的共享语义难定义；使用远程服务时，所有访问被系统串行化，能实现任何集中的共享语义
  - 6) 机器间的接口不同：远程服务方式仅是本地文件系统接口在网络上扩展，季期间的接口是本地客户机和文件系统间的接口；而缓存方式，数据实在服务器和客户机间整体传输，机间接口与用户接口不同
2. 缓存的粒度（数据单元）：可以是文件的若干块，可以是若干文件

- 1) 数据粒度越大，存储的数据就越多，就越增快了访问效率，减少了通信流量和服务器的负载；但传输的一次性开销也会增大，不适用与频繁写的系统
  - 2) 数据粒度小则降低了换成的命中率，增加了通信开销
  - 3) 同理，存储块大小划分也会影响存储性能
3. 缓存的位置
- (1) 客户机和服务器的磁盘、主存，共四个位置
  - 1) 磁盘：可靠，不因系统故障而丢失，提高单个机器的自治性
  - 2) 主存：支持无盘工作站、访问速度快、方便构造单缓存机制（客户机和服务器都用主存换成时，就可使用单缓存，协议简单，易实现）
4. 缓存的更新
- 1) 直接写：修改完立刻写回服务器磁盘，可靠性高
  - 2) 延迟写：先修改到缓存，等段时间再写回，可能（共享）语义不清
  - 3) 驱逐时写：修改过的数据被从缓存中换出时写回
  - 4) 周期性写：周期性扫描缓存，把上次扫描后修改的块写回
  - 5) 关闭时写：关闭文件时写回，对应会话语义
5. 数据一致性
- 1) 客户机发起检查，与服务器联系
  - (1) 检查频度可以是每次访问前进行、打开时进行、周期性进行
  - (2) 检查频率直接影响了网络和服务器的负载
  - 2) 服务器发起检查
  - (1) 需要记录每个客户机缓存的文件
  - (2) 违背了客户/服务器工作模式

## 五. 容错

1. 无状态服务和有状态服务
  - 1) 有状态服务stateful file service：服务器向客户机提供服务时，缓存该客户机的有关信息，该服务器为有状态服务器
  - (1) 服务过程中的崩溃使所有易失性状态丢失，难以恢复
  - 2) 无状态服务stateless file service：不缓存客户机信息，称无状态服务器
  - (1) 系统崩溃后仍能很快重新向客户机提供服务
  - (2) 但不保留客户机信息使请求消息变长，处理过程变久
2. 容错性
- 1) 可恢复性：某文件操作失败或客户端中断操作时，可转换回原来的一致性
  - 2) 坚定性：某存储器崩溃或存储介质损坏时能保证文件完好
  - 3) 可用性：无论何时都能访问，包括及其崩溃和通信失效时
  - (1) 坚定性可用原子性操作保证，坚定性可用设备冗余保证
3. 可用性与文件复制
- 1) 通过对文件在多服务器上独立备份，增强了文件的可用性
  - 2) 还可将文件访问请求分流到各服务器上，平衡负载，避免性能瓶颈
  - 3) 文件复制需要对用户透明，只在底层用标识符区分
  - 4) 文件更新协议复杂，为保证一致性会付出大量开销

- i.
- ii.
- iii.
- iv.
- v.
- vi.
- vii.
- viii.
- ix.
- x.
- xi.
- xii. -----我是底线-----